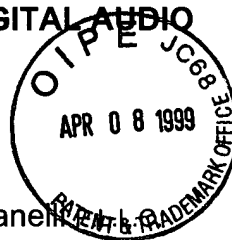


APPLICATION UNDER UNITED STATES PATENT LAWS

Invention: **METHOD AND APPARATUS FOR PROCESSING
INTERAURAL TIME DELAY IN 3D DIGITAL AUDIO**

Inventor(s): Jiashu CHEN; and
Christopher Anton WENDT



RECEIVED
JUN 24 1999
Group 2700

Farkas & Manheim
2000 M Street, N.W.
7th Floor
Washington, D.C. 20036-3307
Attorneys
Telephone: (202) 261-1000

This is a:

- ☐ [] Provisional Application
- ☒ [X] Regular Utility Application
- ☐ [] Continuing Application
- ☐ [] PCT National Phase Application
- ☐ [] Design Application
- ☐ [] Reissue Application
- ☐ [] Plant Application

**SUBSTITUTE
SPECIFICATION**

METHOD AND APPARATUS FOR PROCESSING INTERAURAL TIME DELAY IN 3D DIGITAL AUDIO

This application claims is a continuation of U.S. Patent
5 Application No. 09/191,179 entitled "Method and Apparatus for Regular
Rising Measured HTRF for Smooth 3D Digital Audio" filed November 14,
1998, the specification of which is explicitly incorporated herein by
reference.

10 BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates generally to three dimensional (3D)
sound. More particularly, it relates to a digital implementation of interaural
time delays used in 3D digital sound applications.

15

2. Background of Related Art

Three-dimensional (3D) sound has become integral part of
many personal computer (PC) and consumer electronics devices. It
allows a user to experience realistic sound from any direction using only
20 headphones or speakers.

The rendering of 3D sound involves simulation of a number
of psychoacoustic phenomena occurring when sound is transmitted
through air to each ear. Three of the most important phenomena are
interaural time difference (ITD), interaural intensity difference (IID), and
25 the head related transfer function (HRTF). The ITD is the difference in
time that it takes for a sound wave to reach both ears. The IID is the
sound level difference between each ear. The HRTF is the transfer
function containing any filtering information about the transmission of
sound to a particular ear. This impulse response contains information
30 about the transmission of sound from a particular angular direction,

including any reflections from the shoulder or head and any reflections occurring within the pinna of the ear.

ITD is an important and dominant parameter used in 3D sound rendering. The interaural time difference is responsible for
5 introducing binaural disparities in 3D audio or acoustical displays. In particular, when a sound object moves in a horizontal plane, the interaural time delay is constantly changing depending on the relative location of the sound source and listener. Applying an accurate ITD to a sound can be used to create aural images of sound moving in any desired direction with
10 respect to the listener.

Conventional 3D sound systems embed the interaural time difference in empirically determined HRTFs, typically determined with a mannequin head implanted with microphones in its ears. These delays typically have a relatively large resolution, e.g., 100 microseconds.

15 However, there are at least two basic problems with the implementation of the ITD in a digital environment. In a discrete time environment, time resolution is limited by sampling rate. The traditional use of integer sample delay has limitations. First, the ITD must be rounded to an integer delay, this gives less precision to the rendered ITD
20 delay. Second, a 3D sound rendering which involves motion between multiple angles will incorporate different ITDs. In this situation there will be a discontinuity produced when the renderer switches between each ITD, thus, causing a 'click'. There is thus a need for a method and apparatus for providing a smoothed perceptually 'click-free' 3D sound
25 rendering of the ITD.

SUMMARY OF THE INVENTION

In accordance with the principles of the present invention, a digital delay line for use in a 3D audio sound system comprises a first
30 delay module providing a choice of any delay within the sampling rate

resolution. A second delay module is in series with the first delay module. The second delay module provides a choice of any of a plurality of additional fractional delays.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Features and advantages of the present invention will become apparent to those skilled in the art from the following description with reference to the drawings, in which:

10 Fig. 1 is a block diagram showing the digital 3D sound system including a digital interaural delay line, in accordance with the principles of the present invention.

Fig. 2 is a more detailed diagram showing the digital 3D sound system for creating 3D sound in a digital environment, in accordance with the principles of the present invention.

15 Fig. 3 is a diagram showing the implementation of multiple digital audio streams using a common bank of fractional delay filters, in accordance with the principles of the present invention.

Fig. 4 shows a process for creating an improved ITD look-up table suitable for use in an ITD look up table for use with 3D sound applications as shown in Figs. 1 and 2, in accordance with the principles of the present invention.

Fig. 5 shows a conventional 3D sound system for creating the image of sound from a phantom locality with respect to the listener.

25 Fig. 6 shows a conventional delay line with multiple tap points implemented by Atal-Schroeder.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

In accordance with the principles of the present invention, the ITD is either extracted from measured and empirically determined HRTFs or synthesized using an appropriate head model, smoothed, and

implemented in a look-up table. Implementation of the ITD is provided by a delay line including both an integer portion providing rough estimate delays and a fractional portion providing a very accurate delay and perceptually eliminating discontinuities in the listening field.

5 Fig. 1 is a block diagram showing the basic components of the disclosed embodiment of a digital 3D sound system including a digital interaural time delay line, in accordance with the principles of the present invention.

10 In particular, a sound source **220** is input into a digital interaural time delay line **254**. the interaural delay line **254** includes an integer delay module **250** providing a rough estimate of the desired interaural time delay, and a fractional delay module **252** providing a highly refined additional time delay. In the disclosed embodiment, both the particular settings of both the integer delay module **250** and the fractional
15 delay module **252** are chosen from among a plurality of predetermined delays, greatly reducing or eliminating the otherwise intensive calculations necessary to interpolate a particular interaural time delay.

 The particular delay associated with the left (or right) ear signal **260** and the right (or left) ear signal **262** providing the desired
20 localization of the sound image is provided by a localization control module **270**.

 Fig. 2 is a more detailed diagram showing the digital 3D sound system shown in Fig. 1.

25 In particular, the integer delay module **250** of the disclosed embodiment is comprised of a first-in, first-out (FIFO) buffer **204**. The FIFO buffer **204** may be of any suitable width, e.g., 16 bits, corresponding to the length of the digital audio samples. Moreover, the length of the FIFO buffer **204** will be based on the largest delay necessary to implement the desired 3D sound imaging. The particular delay is related
30 to the selected number of clock cycles after the particular digital audio

sample was input to the FIFO buffer **204**. This selection of an integer delay time is represented in Fig. 2 with a multiplex switch **206**. The use of any of the particular digital audio samples **224a-224d** are fed serially into the FIFO buffer **204**, with the arrows from each of the samples **224a-224d** representing tap numbers.

The clock cycle of the FIFO buffer **204** relates to one over the sample rate. Thus, with an exemplary sample rate of 22 kHz, the 'integer' portion, or resolution of the integer delay module **250** is $1/22,050$ or approximately 45 microseconds (uS).

The second portion of the digital interaural delay line **254** provides a much more refined 'fractional' delay with a fractional delay module **252**. This fractional delay is provided by the selection of any one of a plurality of fractional delay filters **208-212**.

The fractional delay module **252** effectively produces an adjustable digital delay with a finer resolution than the integer delay module **250**. Each of the fractional delay filters **208-212** is a so-called all-pass filter that has a variable phase shift, corresponding to the required fractional delay. The number of phases (i.e., fractional delay filters **208-212**) is determined empirically by behavioral testing of human listening.

In the disclosed embodiment, 64 fractional delay filters are utilized, each providing an incrementally greater delay, in finely resolved increments suitable to the application. For instance, at the exemplary sample rate of 22 kHz, the resolution between the fractional delay filters **208-212** is $(45 \text{ uS})/64$, or about 0.7 uS resolution. This particular fine resolution (and the rough estimate resolution provided by the integer delay module **250**) can be adjusted based on the needs of the particular application.

Each fractional delay filter **208-212** is a finite impulse response (FIR) filter, i.e., a polyphase filter, effecting the desired delay.

Each of the fractional delay filters **208-212**, and/or the fractional delay

controlled switch **216** and/or the multiplexer **214** can be implemented in any suitable processor, e.g., in a digital signal processor (DSP), microprocessor, or microcontroller. Alternatively, the digital filters can be implemented in hardware in accordance with the principles of the present invention.

In the exemplary embodiment utilizing a sampling rate of 22 kHz and 64 fractional delay filters, the first fractional delay filter **208** provides 0.7 uS delay to a digital audio sample, the second fractional delay filter **210** provides approximately 1.4 uS delay, etc., the last fractional delay filter **212** which provides approximately 44.3 uS delay.

Selection of the appropriate fractional delay filter **208-212** is implemented by a multiplexer **214** in the fractional delay module **252**. In the shown embodiment, the fractional delay filters **208-212** are each implemented in a processor, e.g., in a digital signal processor, and selection of an appropriate one of the fractional delay filters **208-212** is desirable at the front end to avoid wasted computational power by running fractional delay filters **208-212** which are not being used for that particular audio sample.

The interaural time delay is controlled by the localization control module **270**, which includes a 3D audio application source position controller **222**, an interaural time delay (ITD) look-up table **220**, and an integral and fractional delay selector **218**. In the disclosed embodiment, the localization control module **270** is implemented in a suitable processor, e.g., in a microprocessor, microcontroller, or digital signal processor (DSP). Of course, the localization control module **270** may alternatively be partially or wholly implemented in hardware, e.g., using programmable array logic.

The 3D audio application source position control **222** selects a desired 'phantom' position of the sound sample currently being input to the digital interaural delay line **254**. The desired location may have a

desired x, y and z coordinate with respect to a reference point, e.g., the center of the listener's head. Based on the desired location, an associated ITD is determined in the ITD look-up table **220**. The integer and fractional delay selector determines the largest integer value which
5 can be achieved within the resolution of the integer delay module **250** without exceeding the desired ITD, and appropriately controls the integer delay module **250** to provide that desired delay to the audio sample. Similarly, the remainder or fractional portion of the desired ITD which is not provided by the integer delay module **250** is provided by an
10 appropriate selection of a desired one of the available fractional delay filters **208-212** in the fractional delay module **252**.

Fig. 3 is a diagram showing the implementation of multiple digital audio streams using a common bank of fractional delay filters, in accordance with the principles of the present invention. Thus, the plurality
15 of fractional delay filters **208-212** can be utilized by a plurality of audio sources for the same listener, avoiding the need to duplicate the fractional delay module **252** for each audio source.

Fig. 4 shows a process for creating the ITD look-up table **220** shown in Fig. 2.

20 In particular, in step **102**, binaural impulse responses are either empirically measured with a sound source at various locations around the listening environment, e.g., at incremental points along a sphere about the sound source or synthesized using an appropriate head model.

25 In step **104**, the ITD information can be extracted from the empirically measured information obtained in step **102**, and a 'mesh' of ITD values for each appropriate point on the sphere is determined. In particular, the ITD samples may be extracted from measured left-right ear head-related transfer functions (HRTFs). These samples can be viewed

as discrete samples of an underline continuous ITD function of azimuth and elevation coordinates.

In step **106**, to avoid undesirable effects for the listener, the ITD mesh determined in step **104** is smoothed using any appropriate smoothing algorithm. For instance, the ITD samples may be regularized using a "generalized spline model" or appropriately filtered and interpolated by a two-dimensional filter to gain smoothness and continuity. While this smoothing may be calculation intensive, it is performed once, off-line, and not performed in real-time as digital audio samples are received.

An ITD mesh can also be synthesized from a head model, i.e. spherical head model, or any other appropriate method of modeling the ITD.

In step **108**, either the smoothed ITD mesh or synthesized ITD samples are input into the ITD look-up table **220**. The ITD mesh may utilize any appropriate coordinate system, e.g., spherical coordinates or a standard x, y and z coordinate system.

In the disclosed embodiment it was determined that the finest time resolution of the overall delay, i.e., the combination of the delay provided by the integer delay module **250** and the fractional delay module **252**, is preferably less than 1 microsecond (μ S) such that any discontinuity caused in the sound stream is under the perceptual threshold of a typical human. In the case of a high sampling rate, faster time resolution may be preferred. For example, with a 22.05 kHz sampling rate of an audio stream, a 64-phase polyphase filterbank was used to obtain sub-microsecond resolution in the time delay.

While the fractional delay filters **208-212** in the disclosed embodiment are each a FIR (polyphase) filter, the principles of the present invention are equally applicable to the use of other filters or digital delays which provide the required delay in a digital audio sample.

The digital interaural delay line **254** in accordance with the principles of the present invention can be implemented in any suitable processor or computer system. For instance, the digital interaural delay line **254** can be implemented at a host level in a personal computer (PC) based platform using regular instruction sets or MMX™ technology, or can be implemented in a digital signal processor (DSP).

To further improve upon efficiency in accordance with the principles of the present invention, the delay may be fixed for one ear, and varied for the sound intended for the other ear, according to the desired movement of the source sound. This alternative method may save as many as half of the instruction cycles required to otherwise process a variably delayed sound to both ears.

The appropriately delayed left and right ear signals can be forwarded to a next stage for further processing, or sent directly to headphones or loudspeakers for presentation to the listener, as a simple binaural signal processing method.

Since ITDs are extracted or synthesized, processed, and implemented separately in a roughly resolved delay module (i.e., the integer delay module **250**), and in a finely tuned delay module (i.e., the fractional delay module **252**), the 3D audio effects can be easily controlled and adjusted to suit other special requirements, e.g., to be optimized for different head sizes. The super resolution sub-sample filtering polyphase filter based delay lines in accordance with the principles of the present invention introduce necessary delay without introducing discontinuity or 'clicks' in the presentation to the listener.

The principles of the present invention are applicable for use in any 3D audio system that uses an interaural time delay as a localization cue for perceived direction of the sound by the listener. For instance, the present invention relates to 3D sound positioning in gaming, virtualizing multiple loudspeaker array systems having two physical

speakers in AC3/Dolby™ Digital systems, advanced computer user interfaces, virtual acoustic reality software for architectural walk-throughs, auralization hardware/software, 3D enhancement for general stereo and wireless headphone sets, etc.

- 5 While the invention has been described with reference to the exemplary embodiments thereof, those skilled in the art will be able to make various modifications to the described embodiments of the invention without departing from the true spirit and scope of the invention.